Crowdsourcing Wildlife Data from Social Media

Saccha Agriel[1], Mahmooda Ali[1], Krishna Vamsi Chndu[1], Kyla Guru[2], Josephine Huss[1], Sourav Jayaprakash[1], Ellen Kidane[1], Viktor Kirillov[1], Shirley Li[1], Varun Mallampati[3], Jared Manusig[1], Hannah Mcdougall[3], Jason Obrycki[3], Michelle Ramirez[1], Zoe Wachtel[4], Tanya Berger-Wolf[1]

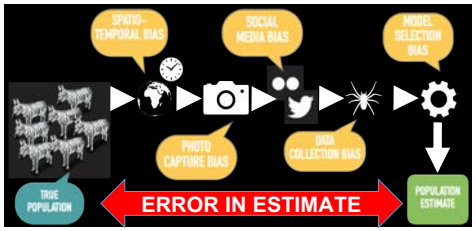
[1]University of Illinois at Chicago, [2]Deerfield High School, [3]Adlai E. Stevenson High School, [4]University of Illinois at Urbana-Champaign

Rationale:

- Traditional wildlife monitoring requires expensive and time consuming data collection.
- Social media is a source of large, cheap, and dynamic data.
- Computer vision and ML (Wildbook™) allow the use of images as the source of information about wildlife.
- Social media data is biased, which affects monitoring results when using social media images as data.

Questions:

- What is the bias of sharing animal images on social media?
- **Does it vary by species and media type?**
- **Does the bias vary by geography and time?**



Our Contributions:

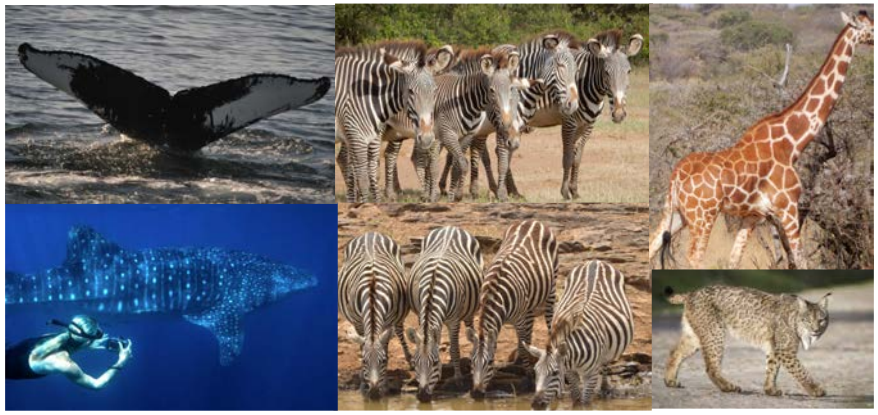
A framework for data collection and the resulting datasets for of images of wild animals posted on social media platforms

Preliminary analysis of wild animal postings by media type and animal species

Platforms:

YouTube flickr iNaturalist

Species:



Process (for each platform for each species):

Use (Python) API + keyword search to find related images → Filter (manually) relevant images. Filter wild/captive → Access metadata → Analyze

Results

Species (color code)	Total Count	Total Relevant %	Total Wild %
Humpback Whales	642	86.4	86.3
Whale Sharks	178	34.3	27.0
Grevy's Zebras	30	70.0	26.7
Plains Zebras	220	74.1	59.5
Reticulated Giraffes	20	80.0	30.0
Iberian (Spanish) Lynx	262	48.9	30.9